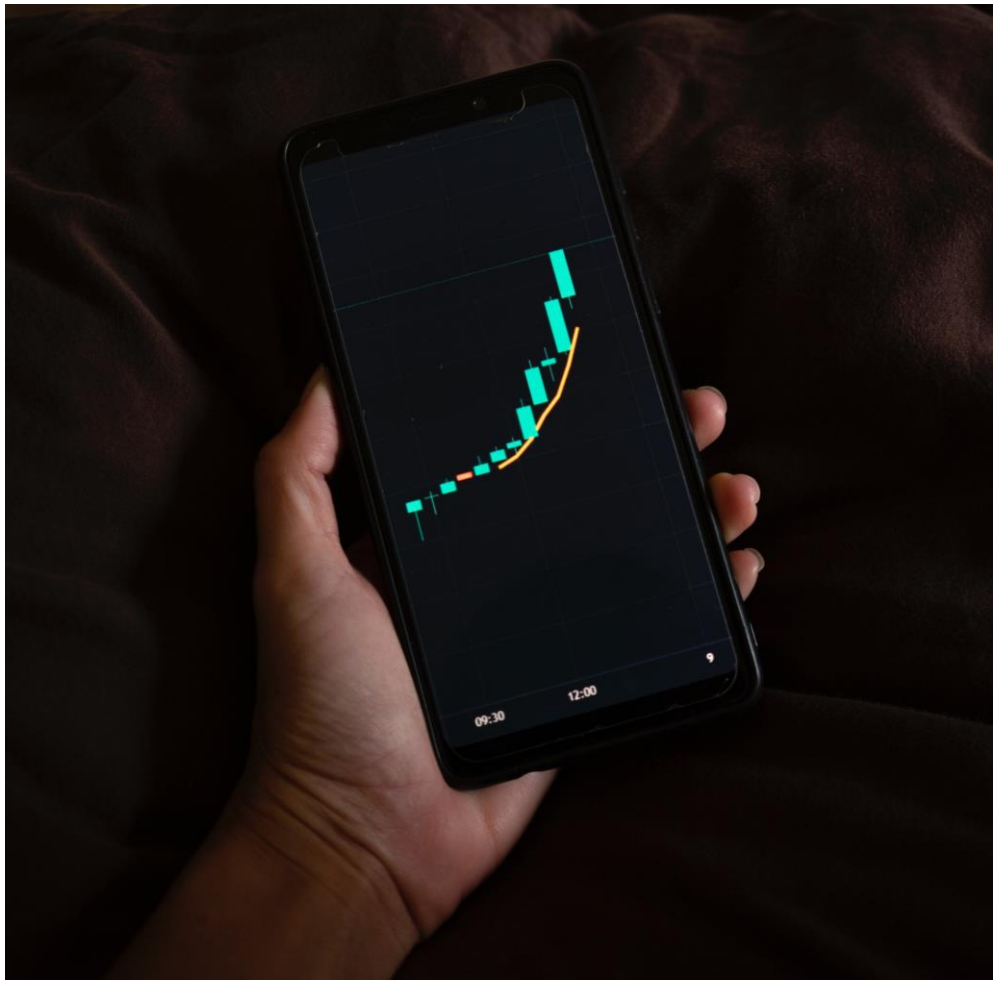
Stock Price Prediction System

SWE3028: Capstone Project

2020312141 Donghun Jung
2019313902 Chanyoung Lee
2018314589 Yujin Seo

Remind Our Project

- Stock Price Prediction System
  - Not real-time
  - Focus on few stocks such as Samsung Electonics
  - Use two models: LSTM, GRU
  - Front-end Implementation



Feedback from Last Presentation

- Verify the prediction system with more diverse stocks
  - Focusing a few stocks is not a thorough prediction
  - It's better to distinguish stocks to predict their prices
    - ✓ ex) stable or unstable

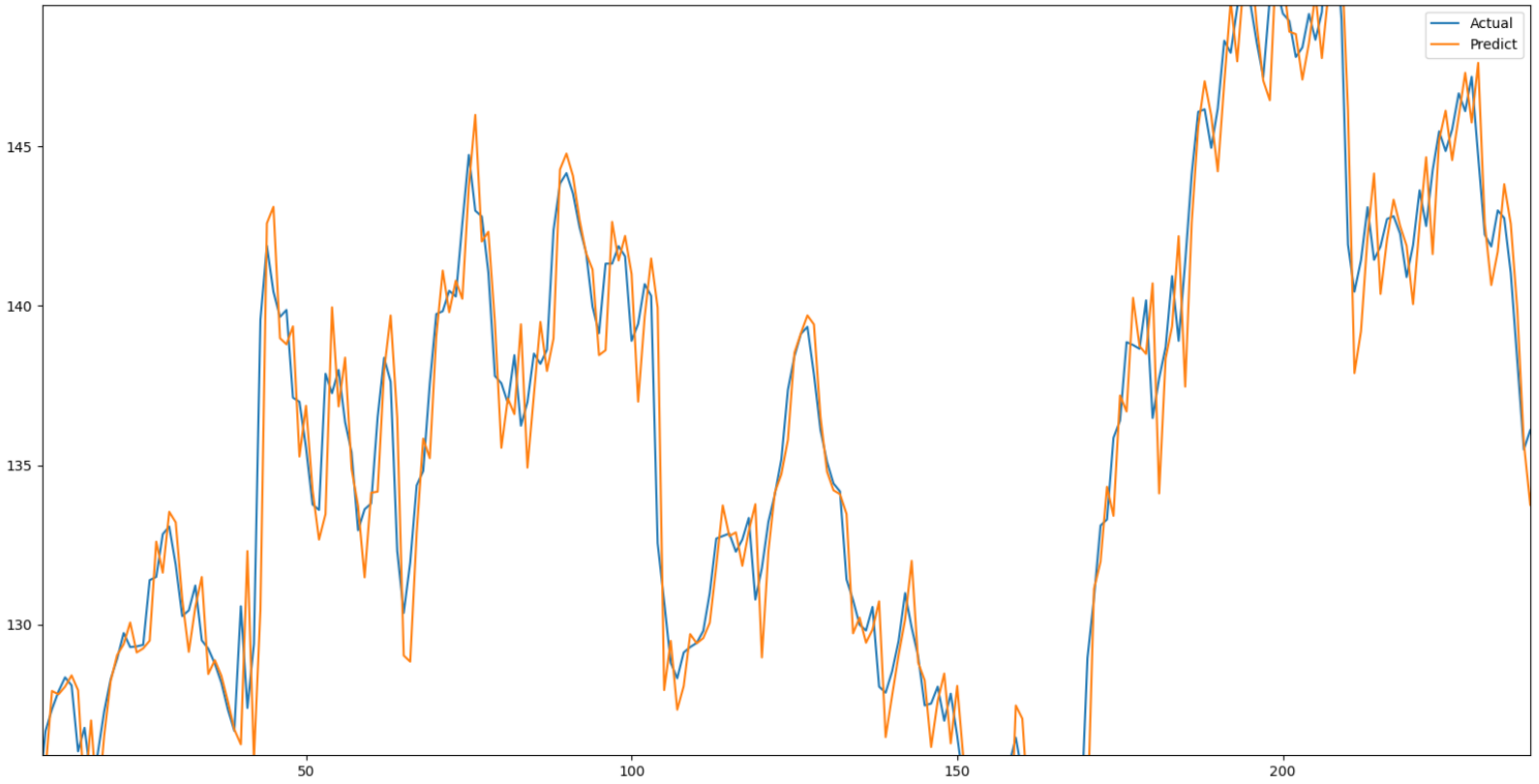
Progress

- Expand the dataset
  - Predict prices of additional stocks (APPL, AMZ, TSLA, NFLX, AMD, etc.)
  - Distinguish stocks into 4 categories


```
us_stable = ['KO', 'MCD', 'WM', 'RSG', 'PEP', 'CL', 'WMT', 'CBOE', 'GD', 'KMB', 'PG', 'COR', 'IBM']

us_unstable = ['APPL', 'MSFT', 'AMZN', 'NVDA', 'META', 'AVGO', 'GOOGL', 'GOOG', 'TSLA', 'ADBE', 'COST', 'CSCO', 'NFLX', 'AMD']

ko_stable = ['005930', '006400', '033780', '000100', '000660', '005830', '010130', '001450', '138040', '030000', '011070', \
             '004170', '024100', '036570', '058470', '011170', '004370', '012750', '081660']
ko_unstable = ['005490', '035420', '005380', '051910', '000270', '068270', '105560', '055550', '373220',  '035720', '012330', \
               '028260',  '207940', '066570', '086790']
```

Progress

- Apply learning rate scheduler to our prediction system



Progress

- Model Implementation: Transformer

```python
class FinanceTransformer(nn.Module):

    def __init__(self, model_args):

        super(FinanceTransformer, self).__init__()

        self.embed_dim = model_args.embed_dim
        self.resolution = model_args.resolution
        self.n_head = model_args.n_head
        self.n_layer = model_args.n_layer
        self.fc_hidden_size = model_args.fc_hidden_size
        self.seq_length = model_args.seq_length
        self.output_length = model_args.output_length

        embed_args = EmbeddingArgs(embed_dim=self.embed_dim, resolution=self.resolution)
        self.embed = FinanceEmbedding(embed_args)

        self.encoder_layer = nn.TransformerEncoderLayer(d_model=self.embed_dim, nhead=self.n_head, batch_first=True)
        self.encoder = nn.TransformerEncoder(encoder_layer = self.encoder_layer, num_layers=self.n_layer)
        self.fc1 = nn.Linear(self.embed_dim*self.seq_length*5, self.fc_hidden_size)
        self.fc2 = nn.Linear(self.fc_hidden_size, self.output_length)
        self.act = nn.ELU()
```

Progress

- Model Implementation: Transformer

```
7  class FinanceTransformer(nn.Module):
```

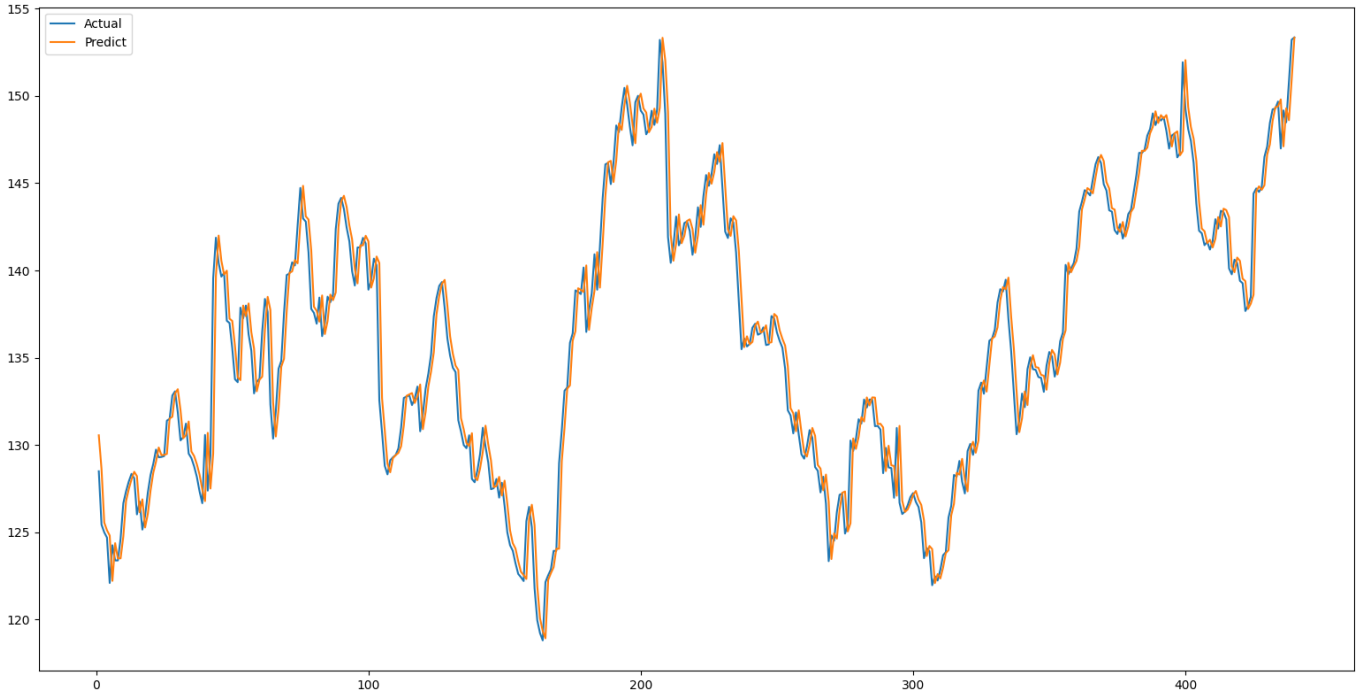
```
30      def forward(self, x):
31
32          embed_x = self.embed(x).reshape(-1,self.seq_length*5,self.embed_dim)
33          enc_x = self.encoder(embed_x).reshape(-1,self.seq_length*5*self.embed_dim)
34          logits = self.fc1(enc_x)
35          logits = self.act(logits)
36          logits = self.fc2(logits)
37          return logits
```

Progress

- Vader Sentiment Analysis
  - Considering recent issues that may affect on stock price
  - Data: Some reports from Securities firms on the stock

Challenges

- Transformer
  - Not



Challenges

- The influence of Vader may be less than expected.
  - The interval between financial reports is not constant.

작성일			
23.11.13	23.11.01	23.10.12	23.09.27
23.11.06	23.10.12	23.10.12	23.09.26
23.11.01	23.10.12	23.10.12	23.09.19
23.11.01	23.10.12	23.10.10	23.09.19
23.11.01	23.10.12	23.10.04	23.09.15

작성일			
2023-11-10	2023-11-01	2023-09-06	2023-06-27
2023-11-08	2023-11-01	2023-07-28	2023-06-15
2023-11-06	2023-10-12	2023-07-28	2023-05-10
2023-11-01	2023-10-12	2023-07-10	2023-04-28
2023-11-01	2023-10-12	2023-07-10	2023-04-28

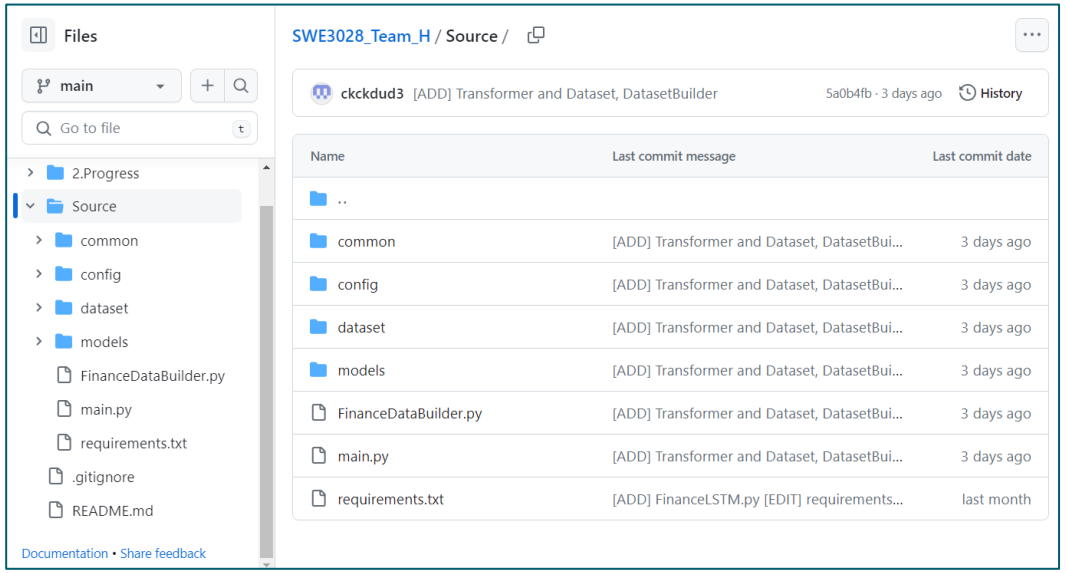
Challenges

- There is a possibility that the score derived from Vader is inaccurate
  - Because the securities report written in Korean has to complete the translation process into English
  - Vader can't score for Korean text

```
[3]  import googletrans
     from googletrans import Translator
     translator = Translator()

[4]  txt = '삼성전자, CES 2024 최고혁신상 3개 수상'
     res = translator.translate(txt, src='ko', dest='en')
     print(res.text)

     Samsung Electronics won 3 CES 2024 Best Innovation Award

     sentiment.polarity_scores(res.text)

     {'neg': 0.0, 'neu': 0.222, 'pos': 0.778, 'compound': 0.9325}
```

Teamwork

- Weekly/Biweekly Meeting
  - Share progress and give feedback to other team members
- Github Code Sharing
  - Share and check our progress



Plan

- Web scraping of securities firms' reports, evaluating them as Vader, and applying them to prediction system
- Select DP-LSTM as baseline for comparing our prediction system
- Update front-end UI/UX design

Schedule

<table>
<tr><th></th><th>Week 13</th><th>Week 14</th><th>Week 15 ~ 16</th></tr>
<tr><td>Donghun Jung
(Front-end)</td><td colspan="2">Update UI/UX Design</td><td rowspan="3">Testing</td></tr>
<tr><td>Chanyoung Lee
(AI)</td><td rowspan="2">Blended Model
&
Vader</td><td rowspan="2">Hyperparameter
Tuning</td></tr>
<tr><td>Yujin Seo
(AI)</td></tr>
</table>